ATTACHMENT C

Copyrighted page(s) have been removed from this document.

Please contact the EPA Docket Center, Public Reading Room to view this document.

Address:
1301 Constitution Ave, NW
Room 3334
Washington, DC 20004

Telephone:
202-566-1744

Fax:
202-566- 9744

Email:
docket-customerservice@epa.gov

ATTACHMENT D

MALCOLM PIRNIE

Technical Memorandum

Date: July 16, 2009

To: Christopher D. Pomeroy, Esq., AquaLaw

From: Clifton F. Bell, P.E., P.G., Malcolm Pirnie, Inc.

Re: Analysis of January-May Inflows to the Chesapeake Bay during the 1996-98 Period

BACKGROUND

Under USEPA guidance (40 CFR 130.7), total maximum daily loads (TMDLs) must be developed to attain water quality standards under critical conditions. For many TMDLs, critical conditions are defined as a hydrologic condition of a given return frequency, such as the 7Q10 streamflow or a storm of a specific return period. For the Chesapeake Bay nutrient TMDL, USEPA plans to model attainment of dissolved oxygen (DO) standards for a ten-year period representing 1991-2000 hydrology. The intention is to meet the critical conditions requirement by basing the TMDL on the "worst" 3-year attainment period within the larger 10-year period.

Preliminary model results indicate that the controlling 3-year period is 1996-1998. In Bay segments such as CB4, attainment of DO standards in 1996-98 is projected to require more nutrient load reductions than for other 3-year periods within the 1991-2000 hydrologic period (CBPO, 2009). A question has arisen regarding whether the 1996-98 period represents unusual hydrologic conditions, or more precisely, whether it represents a hydrologic condition of a longer return period than is normally selected to represent critical conditions for a TMDL. This technical memorandum presents an investigation of that question.

It is well established that the magnitude and extent of hypoxia in the Chesapeake Bay is largely controlled by the magnitude of freshwater and nutrient inputs during the preceding winter and spring months (Malone and others, 1993; Boesch and others, 2001). Freshwater input during this period affects the extent of hypoxia not only by conveying a large proportion of the annual nonpoint source nutrient loads, but also by affecting the degree of stratification of the Bay water column. Scavia and others (2006) developed a simple empirical model of Bay hypoxia as a function of nutrient inputs from January to May, and this model is now used annually to forecast the size of the "dead zone" that develops in late spring and summer. The amount of freshwater inflow to the Bay during January-May, therefore, is a useful indicator of hydrologic conditions associated with DO standards attainment.

METHODS

The daily average input of freshwater flow to the Chesapeake Bay was computed as the sum of daily average streamflows at two USGS stream gaging stations:

- Susquehanna River at Conowingo Dam (USGS 1578310); period of record: Oct 1967 to June 2009
- Potomac River near Washington DC (USGS 1646502); period of record: March 1930 to May 2009

The total inflow to the Bay will be higher than the sum of the inflow at these two stations. However, flows from the Susquehanna and Potomac Rivers together represent almost 80 percent of the gaged inflows to the Bay (Sprague and others 2000), and an even higher proportion of gaged inflows that strongly affect hypoxia in the critical mid-Bay segments. The overlapping period of record for these stations was October 1967 to May 2009, a period of about 42 years. The average daily inflow from January through May was calculated for each year in this period. The average daily inflow from January through May was also calculated for each of the forty 3-year periods within the 42-year period.

RESULTS AND DISCUSSION

Results (Table 1) demonstrate that the 1996-1998 period had the highest average Jan-May inflow of the entire period of record, representing the 100th percentile of the data. Because this period represents one of forty 3-periods included in the analysis, the resulting estimate of return period is 40 years.

The 1996-1998 period is so usual because it contains two years—1996 and 1998—that represent the 93rd and 98th percentiles, respectively, of Jan-May inflows. Although it is not extremely rare for any given 3-year period to have one such year, it is rare for any 3-year period to have two such years. High inflows in the year 1996 are partly due to extreme meteorological/hydrologic conditions. In January 1996, warm rains fell on a winter snowpack and caused an event known as the "Big Melt". This event has been labeled an "extreme" event by the Chesapeake Bay Program Office and required special consideration during calibration of the Chesapeake Bay simulation models (Shenk, 2008). Inflows during January-May of 1998 were even higher than in 1996.

USEPA guidance does not define "critical conditions" nor address the issue of reasonable return periods for TMDL development. However, a survey of nationwide TMDL documents reveal that the vast majority of TMDLs are developed for hydrologic conditions that represent return periods of 10 or fewer years. The majority of TMDLs developed for critical low flow conditions have used the 10-year return period associated with 7Q10 or 1Q10 streamflow statistics. The reviewed identified TMDLs developed for high flow conditions that used specific design storms with return frequencies of 1, 2, 5, or 10 years. Based on this non-comprehensive review, no specific TMDL examples were discovered that used a return period of 40 years or higher, although some might exist.

Based on this analysis, the critical condition currently being planned for the Chesapeake Bay TMDL appears to be significantly more infrequent than is normally used for TMDL development. Flow percentiles such as those presented in Table 1 can be used to select alternate 3-year periods that represent critical but not extreme conditions. For example, the 1993-1995 and 1994-1996 periods had very high January-May inflows, but were much closer to a 10-year return period than the 1996-1998 period.

REFERENCES

Boesch, D. F., R. B. Brinsfeld, and R. E. Magnien. 2001. Chesapeake Bay eutrophication: Scientific understanding, ecosystem restoration, and challenges for agriculture. *Journal of Environmental Quality* 30:303-320.

Chesapeake Bay Program Office, 2009, *Basinwide Cap Load Analysis: Findings and Recommendations*. Attachment A of materials for the Chesapeake Bay Program Water Quality Steering Committee conference call on June 22, 2009. 20 p.

Malone, T. C., W. Boynton, T. Horton, and C. Stevenson. 1993. Nutrient loading to surface waters: Chesapeake case study, p 8-38. In M. F. Uman (ed.), Keeping pace with science and engineering. National Academy Press, Washington, DC.

Scavia, D., E.A. Kelly, and J. D. Hagy III. 2006. A simple model for forecasting the effects of nitrogen loads on Chesapeake Bay hypoxia. Estuaries and Coasts 29(4): 674-684.

Shenk, G. 2008. Hydrologic average period. Presentation delivered to the Chesapeake Bay Program Water Quality Steering Committee on April 22, 2008. 26 p.

Smith, D. E., M. Leffler and G. Mackiernan (eds.). 1992. *Oxygen Dynamics in the Chesapeake Bay: a Synthesis of Recent Research*. Pp. 61-112. Maryland Sea Grant College. UM-SG-TS-92-01. College Park, Maryland.

Sprague, L.A., Langland, M.J., Yochum, S.E., Edwards, R.E., Blomquist, J.D., Phillips, S.W., Shenk, G.W., and Preston, S.D. 2000. Factors Affecting Nutrient Trends in Major Rivers of the Chesapeake Bay Watershed. USGS Water-Resources Investigations Report 00-4218. 109 p.

cfb

MALCOLM PIRNIE

Technical Memorandum

Date: September 15, 2009

To: Gary Shenk, USEPA Chesapeake Bay Program Office

From: Clifton F. Bell, Malcolm Pirnie, Inc.

Re: Evaluation of Monthly Span for Critical Hydrologic Period

It is our understanding that the CBPO is proceeding with additional analyses of the critical hydrologic period issue, following up on discussion of the WQGIT teleconference of September 9, 2009. One of the technical issues discussed on that call was that of the monthly span for defining critical hydrologic conditions. This technical memorandum presents an evaluation of this issue with recommendations for consideration by the CBPO as they proceed with their analysis.

BACKGROUND

Malcolm Pirnie had originally used a January-May span for the hydrologic analysis, based on Bay-related scientific literature that either explicitly used this period in statistical modeling of Bay hypoxia or otherwise emphasized the importance of the winter-spring freshet in not just delivering loads but also strengthening stratification and setting a starting point for D.O. decline (*e.g.*, Hagy and others, 2004; Scavia and others, 2006; Stow and Scavia, 2008; Seliger and Boggs, 1988; Boicourt, 1992; Boynton and Kemp, 2000). Preliminary analysis by Tetra Tech, as presented on the September 9 WQGIT call, demonstrated that the average monthly stream flow of longer monthly spans (e.g., September-June) had higher R^2 values when regressed against DO violations rates. Return periods of critical hydrologic conditions can be expected to be sensitive to the monthly span chosen for averaging. Therefore, it is important to determine what monthly span is the most statistically and mechanistically appropriate for defining critical conditions.

To assist with this evaluation, Malcolm Pirnie performed the following: (1) contacted Dr. James Hagy of the USEPA to determine the basis for the January-May span used in the Bay "dead zone" forecasting model; (2) investigated why inclusion of stream flow from the previous fall might give increased R^2 values; and (3) evaluated alternative (non-parametric) means for quantifying the correlation between Bay inflows and DO violation rates. Based on this analysis, we recommend that monthly span start in either December or January and end in either May or June. To address uncertainty with the appropriate monthly span, return periods could be expressed as ranges associated with the four possible monthly spans.

ORIGIN OF JANUARY-MAY PERIOD

The January-May period is used in a well-known statistical model to predict hypoxia as a function of winter-spring nutrients loads to the Bay (*e.g.*, Hagy and others, 2004; Scavia and others, 2006; Stow and Scavia, 2008). The model had its origin in Ph.D. dissertation work by James Hagy. Malcolm Pirnie communicated with Dr. Hagy on September 10, 2009, and inquired about the basis for the January-May period. Dr. Hagy's response can be paraphrased as follows:

- There is nothing binding about the January-May period specifically; it happened to provide the best prediction of hypoxia for the dataset with which he was working.
- However, mechanistically speaking, it is the winter-spring freshet that is of most interest in determining the potential for summer hypoxia.
- Streamflows as far back the previous September are not expected to have a significant mechanistic-hydrologic effect on summer hypoxia. Higher correlations with such as longer period are probably due to chance.
- The timing of the freshet varies from year to year. It most often occurs in the late winter or spring (Mar-May), in some years, streamflows as early as December can affect the salinity regime and the potential for hypoxia.
- The Jan-May period tends to capture the months that are most often important, although not all these months might be important in any given year.

EFFECT OF HIGH LEVERAGE DATA ON R^2 VALUES

The lack of a strong mechanistic basis for the effect of early fall streamflows on summer hypoxia leads to the question of why the inclusions of these months might increase R^2 values of the streamflow-hypoxia regression. The addition of December to the monthly span might actually improve the mechanistic basis of the relation, because in some years December flows might be an important component of the winter-spring freshet, as discussed above. However, R^2 value can be very sensitive to individual data points of high leverage, particularly in relatively small datasets such those with which we are dealing. This seems to be the case with the inflow-DO violation rate regression.

Figure 1 is a scatter plot of the 3-year average Bay inflow (Potomac + Susquehanna) v. 3-year DO violation rate developed using 1985-2006 data. Bay inflow averages were computed both as Jan-May and Sep-June averages. The slopes of the two relations are almost identical, and both are highly significant regressions. The September-June regression has a slightly higher R^2 value. However, when the single data point associate with the highest DO violation rate (associated with 2003-2005) is removed, the two R^2 values are identical (Figure 2). Given the sensitivity of R^2 values to individual data of high leverage, we believe that it would be useful to examine the correlations between average streamflow and DO violation would best be examined using non-parametric statistics.

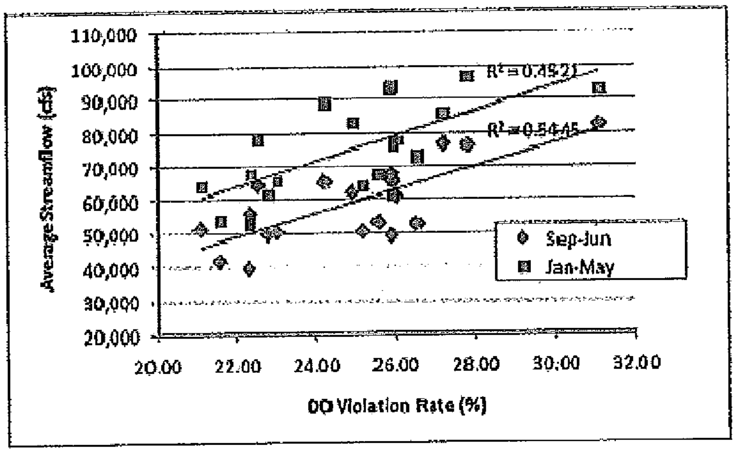


Figure 1: Scatterplots of 3-year average Bay inflows v. 3-year DO violations rates, 1986-2006.

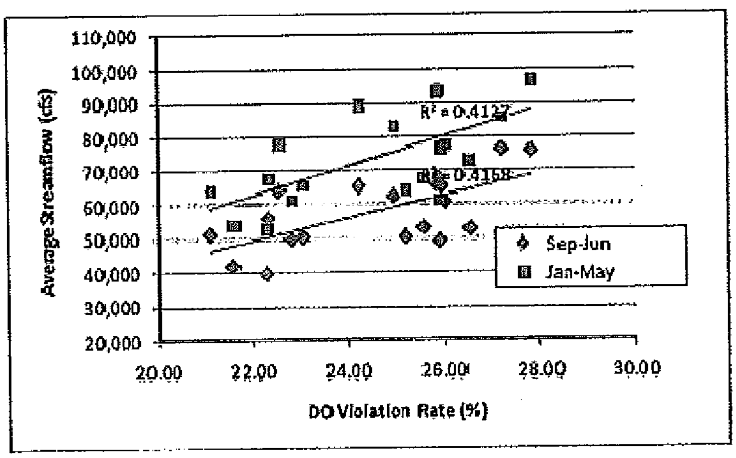


Figure 2: Scatterplots of 3-year average Bay inflows v. 3-year DO violations rates, 1986-2006, with 2005-2006 datum removed.

NON-PARAMETRIC CORRELATIONS BETWEEN DO VIOLATION RATES AND AVERAGE STREAMFLOW

Correlations between DO violation rates and average Bay inflows were computed using two non-parametric statistics: Spearman's rank correlation coefficient, and Kendall's tau. Different monthly spans were used to compute the average Bay inflow, the longest period being January-June and the shortest period being January-May. Results (Table 1) demonstrate that all the correlations are highly significant and in a similar range (0.6-0.7 for Spearman's rank correlation coefficient and 0.4-0.5 for Kendall's tau). The addition of September-November to the inflow did not increase the correlations, and in fact decreased them slightly. The addition of December and January to the January-May period increased the correlations slightly.

TABLE 1
Non-Parametric Correlation Coefficients for 3-Year Average Bay Inflows v. 3-Year D.O. Violation Rates

Monthly Span for Inflow Average	Spearman R	p-level	Kendall Tau	p-level
Sep-Jun	0.64	<0.01	0.48	<0.01
Oct-Jun	0.66	<0.01	0.51	<0.01
Nov-Jun	0.70	<0.01	0.52	<0.01
Dec-Jun	0.72	<0.01	0.56	<0.01
Jan-Jun	0.67	<0.01	0.47	<0.01
Sep-May	0.61	<0.01	0.45	<0.01
Oct-May	0.61	<0.01	0.46	<0.01
Nov-May	0.66	<0.01	0.49	<0.01
Dec-May	0.60	<0.01	0.44	<0.01
Jan-May	0.61	<0.01	0.44	<0.01

RECOMMENDATIONS ON MONTHLY SPAN

Based on this analysis, we do not believe the September-November streamflow adds mechanistic information to the analysis, and thus we recommend that the monthly span for the hydrologic analysis remain representative of the winter-spring freshet, without addition of early fall inflows. Given the similar correlation coefficients for different monthly spans, and the relatively small data set for such computations, one should not choose between them on the basis of correlation coefficients alone. The January-May period remains of interest due to the fact that it captures the months that are most often important, and the use of this period has a strong precedent in the Bay hypoxia forecasting model.

The addition of December or June to the monthly span could also be considered. Given that calculated return periods could be sensitive to the monthly spans chosen, one manner to proceed would be to calculate the returned intervals associated with 2-4 of the primary spans of potential interest (Jan-May, Jan-Jun, Dec-May, Dec-Jun) and express the return periods as a range.

REFERENCES

Boicourt, W. C. 1992. Influences of circulation processes on dissolved oxygen in the Chesapeake Bay, p. 7–59.In D. E. Smith, M. Leffler, and G. Mackiernan (eds.), Oxygen Dynamics in the Chesapeake Bay. A Synthesis of Recent Research. Maryland Sea Grant College. College Park, Maryland.

Boynton, W.R. and W.M. Kemp. 2000. Influence of river flow and nutrient loading on selected ecosystem processes: a synthesis of Chesapeake Bay data. pp. 269-298, In: Hobbie, J. (Ed.), A Blueprint for Estuarine Synthesis. Beckman Center, Univ. of California at Irvine, CA. [UMCES Contribution No. 3224-CBL].

Hagy, J.D., W.R. Boynton, C.W. Keefe, and K.V. Wood. 2004. Hypoxia in Chesapeake Bay, 1950–2001: long-term change in relation to nutrient loading and river flow. Estuaries 27:634-658.

Scavia, D., E.A. Kelly, and J. D. Hagy III. 2006 A simple model for forecasting the effects of nitrogen loads on Chesapeake Bay hypoxia . Estuaries and Coasts 29(4) 674-684.

Stow, C.A., and D. Scavia. 2008. Modeling hypoxia in the Chesapeake Bay: Ensemble estimation using a Bayesian hierarchical model. J. Marine Systems, 76:244-2.

ATTACHMENT E

MALCOLM PIRNIE

Technical Memorandum

Date: May 27, 2010

To: Virginia Association of Municipal Wastewater Agencies
Maryland Association of Municipal Wastewater Agencies

From: Clifton F. Bell

Re: Recommendations on Baywide Loading Targets

On the May 24, 2010 Water Quality Goal Implementation Team (WQGIT) teleconference, USEPA presented the latest model results of dissolved oxygen (DO) and chlorophyll-*a* attainment under various loading scenarios. The USEPA announced its intention to derive the initial Baywide cap for nitrogen and phosphorous in the next week. At the conclusion of the teleconference, USEPA asked the states to provide quick feedback (1-2 days) via email on the appropriate Baywide target for main stem DO standards attainment.

The purpose of this memo is to address the present Baywide load allocation question. Highlights of this technical review are as follows:

- Given the high cost of management actions, it would be recommended to adopt an allocation approach that recognizes the proper uses and limitations of the modeling framework, and thus avoids:
    - Large swings in allocations between model versions.
    - Large swings in allocations to achieve numerically insignificant increases in attainment rates.
- The Bay Program modeling results should not be used in a manner that overestimates the precision of the model. The load-response predictions should be examined for asymptotic relations that would cause the target loads to be highly sensitive to small changes in nonattainment that exceed the precision of the model. In these cases, a difference in DO percent non-attainment rate of 3-5% should be used as a general guide to establish which model scenarios are "essentially equivalent."
- Based on the guideline cited above, the present Baywide nutrient load target should be based on the Target Load Option A scenario (200 Mlb/yr TN, 15 Mlb/yr TP). This loading recommendation was previously presented to and approved by the PSC.
- Due to modeling problems in shallow open water segments, as well as in embayments, the present model should not be used to adjust allocations for smaller local segments. At this time, the Baywide allocation process should be limited to deep

water and deep channel DO on the larger mainstem Bay problem segments including CB4MH, CB5MH, MD5MH, and VA5MH (as previously done).

- The Baywide allocation should not be based on side/tributary segments such as CHSMH, MAGMH, EASMH, or segments of the Elizabeth River, which can experience sensitivity local controls, natural causes of nonattainment, or local model calibration/resolution issues. In many cases, allocating to these smaller problem segments could require local modeling refinements. A phased approach is recommended to effectively address remaining problem segments.

- Both the magnitude and location of loads must remain primary considerations in the derivation of basin wide target loads. A further reduction of in loads from the southern tributaries (i.e., the James and York) would not significantly influence mainstem Bay DO attainment.

These specific recommendations are discussed in more detail below:

1. *The allocation approach should recognize the proper use and limitations of the modeling framework.* The Chesapeake Bay Program's framework of linked models, while very sophisticated, is still only an approximation of the natural system. The models were originally intended to provide an approximation of the large-scale hypoxic volumes under various loading scenarios. Under the present TMDL process, the model output is now being interpreted at spatial and temporal scales that exceed its precision. The ability of the model to distinguish small differences in attainment rates between model scenarios should be questioned considering factors such as:

- Continued instability in predicted attainment rates with each new version of the model.
- The documented ability of small numbers of outliers in the observed data set to cause predictions of non-attainment.
- The lack of a validation documenting the reliability of the model to accurately predict response to large-scale loading reductions that are simulated in the scenarios.

The model framework as a whole is conservative due to various assumptions such as:

- All point sources discharging maximum loads at all times.
- Selection of many conservative BMP efficiencies (as previously commented on by V/MAMWA).
- Allocations based on very small regions of the Bay system, which are much lower than needed for the great majority of the system.

Model uncertainty will be addressed by an implicit of margin of safety associated with the conservativeness of the model. But in the context of choosing a Baywide load allocation, the primary question becomes the following: *What is the ability of the model (or lack thereof) to truly differentiate attainment rates between scenarios?*

Consider the following hypothetical illustration: If Scenarios A, B, and C produce non-attainment rates of 25%, 6%, and 4%, one might believe that Scenarios B and C offer significant improvements from Scenario A. However, in reality Scenarios B and C are themselves essentially equivalent in terms of their response. It would be improper to make large cuts in allocations—with huge cost implications—on the basis of such small numerical differences in predicted attainment rates between B and C given the true sensitivity of the model.

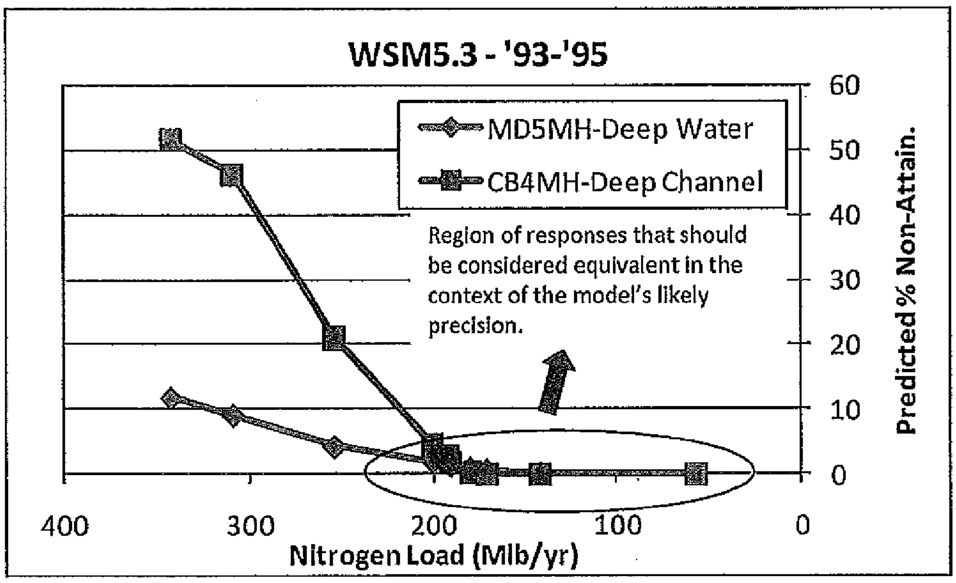


Figure 1: Example of asymptotic relations between predicted loading and non-attainment rates. Anywhere along the flat part of the curve, load allocations will be highly sensitive to very small changes in predicted non-attainment. These changes exceed the likely precision of the model.

2. <u>*The target load option A (200 Mlb/yr TN, 15 Mlb/yr TP) represents the load allocation at which key model segments are predicted to be in attainment with DO standards*</u>. Based on an examination of the most recent stoplight plots, the tributary strategy (200/15) appears be the scenario at which key model segments are predicted to come into attainment, and/or the scenario <u>beyond which more stringent scenarios are essentially equivalent in their water quality result (although vastly more stringent in terms of the associated management measures necessary to achieve that water quality result)</u>. The following is an examination of the key mainstem Bay segments (CB4MH and CB5MH) under the target load option A scenario:

- CB4MH
  - Deep water: In attainment
  - Deep channel: Within ~2% of attainment; equivalent to more stringent loading scenarios.
- CB5MH (Entire segment)
  - Deep water: In attainment
  - Deep channel: In attainment
- MD5MH (Maryland-only portion of CB5)
  - Deep water: Within ~2% of attainment; equivalent to more stringent loading scenarios.
  - Deep channel: In attainment
- VA5MH (Virginia-only portion of CB5)
  - Deep water: In attainment.
  - Deep channel: In attainment.

Considering the model limitations, it cannot be concluded that allocations lower than the target load option A would significantly improve attainment rates. Therefore, Target Load Option A is the most appropriate basis for the next Baywide loading target.

This loading scenario would not immediately address attainment of side segments including the CHSMH, MAGMH, and EASMH segments for deep water and deep channel DO. However, a comparison between results obtained by WSM5.1 and WSM5.3 indicate wide swings in response to attainment loading rates for these particular segments (Figure 2). Such wide swings indicate that further examination and explanation is needed to understand whether the model's predictions are scientifically defensible as a basis for decision making. Effectively addressing these segments might require separate, locally-oriented modeling analysis with a modeling tool better adapted to evaluating local conditions.

3. *Due to open water modeling issues, the present model should not be used to adjust allocations for local segment open water DO attainment.* The recent work by the Bay Program has highlighted serious model limitations in predicting attainment of open water DO standards. These include mechanistic errors in the simulation of DO in cells adjacent to shorelines, model grid resolution problems in small channels, bias in attainment rates due to a small number of unusual DO observations, and extrapolation of DO concentrations beyond the observed range.

Although mechanistic modeling problems are obvious in some segments due to unexpected load-response relationships, it should be stressed that the same mechanistic modeling problems are likely occurring in many other segments. Until and unless the model is shown to be accurate at simulating open water DO, and the level of accuracy is known, we recommend that these results not be used to further adjust target loads in a downward direction.

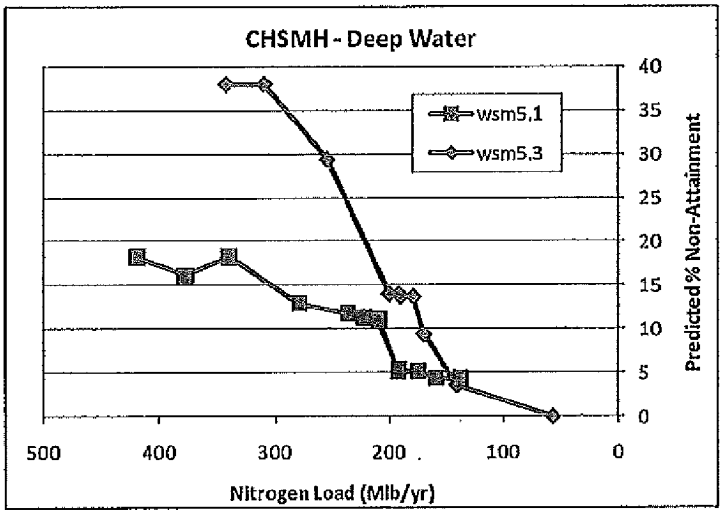


Figure 2: Example of widely different model predictions of non-attainment between model version 5.1 and 5.3 for a tributary segment.

4. *Both the magnitude and location of loads must remain primary considerations in the derivation of target loads*. The southern tributaries (i.e., the James and York) have very little effect on mainstem Bay attainment. Therefore, allocations in these basins should not be adjusted downwards to achieve a Baywide loading cap lower than 200/15. Although such adjustments might help achieve a given magnitude of loading, it would not achieve commensurate water quality benefits. Any adjustments to the magnitude of the Baywide loading cap should explicitly continue to consider the geography of load reductions.

5. *The Chesapeake Bay Program should achieve and communicate a clear understanding of the reasons for instability in predicted attainment rates between model version 5.1 and 5.3*. Based upon the premise that the water quality and sediment transport model (WQSTM) required little to no recalibration for use with watershed model (WSM) version 5.3, in comparison with WSM version 5.1, it is unclear why the different model versions would predict very different nonattainment rates at a given loading level for some segments (e.g., CHSMH, EASMH). The answer to this question is central to understanding whether the variation in predicted attainment rates is associated with manageable variables (e.g., the geography of load reductions) versus non-manageable variables (e.g., differences in the model algorithms). It would also help better quantify the amount of nonattainment that the model can truly distinguish between model scenarios. The Bay Program should diagnose the causes of the differences in model predictions, and clearly communicate these differences to the Bay partners before basinwide targets are selected.

cfb

MALCOLM PIRNIE

Technical Memorandum

Date: June 10, 2010

To: Virginia and Maryland Associations of Municipal Wastewater Agencies

From: Clifton F. Bell

Re: Magnitude of Significant Differences in DO Criteria Violation Rates

This technical memo presents the results of a power analysis to evaluate the minimum difference in dissolved oxygen (DO) violation rates that would be statistically detectable. Results indicated that differences in spatial violations rates of less than about 4-6% would not be measureable even over long monitoring periods. The conservative value might increase using other methods that consider intra-assessment period variability. Other parameters, such as chlorophyll-*a*, are much more temporally and spatially variable than DO, and thus could have significant higher MSD values.

BACKGROUND

The Chesapeake Bay Program's modeling framework is a tool to estimate what improvement in environmental conditions would result if certain nutrient management actions (i.e., scenarios) were put on the ground. Recent discussions with the Chesapeake Bay Program's (CBP) Water Quality Goal Implementation Team (WQGIT) have raised the question of whether the Chesapeake Bay modeling framework (and associated post-processing steps) can differentiate between model scenarios varying little in non-attainment rates. V/MAMWA raised this issue as part of a recommendation to avoid large swings in load allocations based model scenario predictions that are essentially equivalent given the likely precision of the model predictions (Bell, 2010).

The CBP's modeling framework is largely deterministic rather than stochastic, meaning that the predictions are based primarily on physical laws without explicit consideration of randomness or statistical variation. As a result, the actual precision of the non-attainment predictions for future management scenarios cannot be easily quantified. A related question that can be directly addressed is, "What is the minimum difference in non-attainment between two monitoring datasets that can be detected given data variability?" The question is relevant to the target load selection process because the public will have an expectation that the effect of the required controls on the environment be measurable and cost-effective, and water quality monitoring program data will be used to assess these improvements. Many stakeholders would agree that target loads should not be based on very small theoretical differences that exist between scenarios in the "model world" that would not translate to measureable improvements in the "real world".

Statistical power analysis represents a method to determine the magnitude of changes in water quality data that are statistically discernable from the variability that is inherent to the data. This technical memo presents one simple approach to determining the minimum significant differences in non-attainment rates of DO criteria.

POWER ANALYSIS

With a specific segment, month-to-month spatial violation rates are not normally distributed. However, the mean spatial violation rates for different 3-year assessment periods are approximately normally distributed, especially for segments that experience relatively low mean spatial violation rates (<10%). If the mean spatial violation rates for a segment were approximately normally distributed, the minimum significant difference (MSD) in the mean spatial violation rate could be determined by a parametric power analysis. This MSD of the spatial non-attainment rates would provide insight into the MSD of the area under the CFD assessment curve, because that area is calculated as a multiple of the spatial violation rates.

The MSD of spatial violations would be a direct function of the sample size (n), the Type I error rate (α), the Type II error rate (β), or power ($1-\beta$), and standard deviation of the spatial violation rates. In this example, α and $1-\beta$ were set to the conventional values of 0.05 and 0.8, respectively.

In the present example, the sample size n represents the number of three-year periods for which monitoring data are available before and after some treatment, such as the adoption of management practices. The MSD is inversely related to n, such that smaller differences in mean violation rates could be detected over longer monitoring periods. For the purposes of this exercise, n was set to 9, which corresponds to the number of independent (i.e., non-overlapping) three-year assessment periods over a twenty seven-year monitoring period. This approximately corresponds to the pre-TMDL period for which adequate monitoring data are available to assess spatial violation rates (early to mid 1980s – 2010). Hence, the power analysis will approximate the MSD that could be detected between pre-TMDL and post-TMDL monitoring periods up to about 2037.

In general, the standard deviation of spatial violation rates decreases as segments approach overall attainment relative to the CFD curve. For this example, it was desired to use a conservatively low standard deviation, to avoid overestimating the MSD. The selected values (3% and 4%) are typical of standard deviation of the mean spatial violation rates for deep water segments that are in overall attainment with the deep water CFD curve, as determined from a tabulation previously provided by the CBP (Attachment A).

The evaluation was conducted as a power analysis of a two-sample t-test using the software of Lenth (2010). Results indicate, under the assumption of this exercise, the MSD of the mean spatial violation rates is in the range of 4-6% (Table 1). In other words, under the assumptions specified for this analysis, the means of two independent tabulations of mean spatial violation rates would have to differ by 4-6% before they could be determined to be statistically different, even over long monitoring periods. The actual difference in magnitude of overall (time-space) non-attainment rates would depend on the positions and shapes of the segment's curves relative to the respective reference curves. However, the differences in the mean percent area under their respective CFD curves would be highly correlated with the differences in spatial violation rates.

TABLE 1
Power Analysis: Two Sample t-test

Parameter	Value
Type I error rate (α)	0.05
Type II error rate (β)	0.8
Sample size (n); identical for both samples	9
Standard deviation in mean spatial violation rate	0.03 – 0.04
MSD	0.04 – 0.06

DISCUSSION

Based on the results of this analysis, segments that are in or close to attainment would have to have spatial DO violation rates that differ by 4-6% or more before they could be statistically distinguished from one another. Because the power analysis was conducted on the means of violations rates for three-year periods, the analysis did not consider variability of violation rates within three-year periods, which could increase the MSD. It would be recommended to explore other power analysis methods that consider intra-assessment period variation, such as analysis of variance (ANOVA) methods or methods that match observations by the month of measurement. However, the present analysis is analogous to the current assessment methodology by which a single non-attainment rate is estimated for each three-year period, without explicit consideration of the uncertainty of that value.

Results of the power analysis have indirect rather than direct bearing on the question of the precision of model non-attainment predictions. The model's precision is related to a host of factors other than the variability in the monitoring data, including the resolution of input datasets, calibration, variability associated with regressions developed for model post-processing, and variability between model versions. However, the power analysis does demonstrate that small (<4-6%) differences in model predictions of attainment between scenarios would probably not be measureable in the "real world".

This analysis is most pertinent to predictions of DO attainment in mainstem Bay segments. Other parameters, such as chlorophyll-*a*, are much more temporally and spatially variable than DO, and thus could have significant higher MSD values.

REFERENCES

Bell, C.F. 2010. Recommendations on Baywide Loading Targets. Technical memorandum delivered to the Virginia and Maryland Associations of Wastewater Agencies. May 27, 2010. 5 p.

Lenth, R. V. 2010. Java Applets for Power and Sample Size [Computer software]. Retrieved June 4, 2010, from http://www.stat.uiowa.edu/~rlenth/Power.

cfb

ATTACHMENT A

Spatial Violation Rates of Attaining Deepwater Segments

[Data from elec. comm., Excel file entitled "DO_violation_rates.xls", provided by J. Keisman to J. Pletl on 14 May 2009]

Segment	CB6PH	CB6PH	CB6PH	CB6PH	CB7PH	CB7PH	CB7PH	CB7PH	CB7PH	CB7PH	YRKPH
Period	96-98	97-99	99-01	04-06	96-98	97-99	98-00	99-01	00-02	04-06	04-06
Ranked Monthly Spatial Violations Rates - Deep Water 30-Day Mean Criterion	40%	51%	51%	40%	40%	40%	40%	26%	3%	23%	39%
	20%	40%	19%	27%	9%	26%	26%	3%	3%	21%	39%
	16%	20%	7%	25%	5%	9%	5%	3%	1%	5%	25%
	11%	9%	2%	20%	4%	5%	3%	0%	0%	3%	11%
	9%	2%	2%	9%	2%	1%	3%	0%	0%	1%	4%
	8%	0%	0%	3%	1%	1%	1%	0%	0%	1%	0%
	2%	0%	0%	3%	1%	0%	0%	0%	0%	0%	0%
	0%	0%	0%	3%	1%	0%	0%	0%	0%	0%	0%
	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%
	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
3 YR MEAN	9%	10%	7%	11%	5%	7%	7%	3%	1%	5%	10%

Note: 3 YR means are approximately normal with mean= 7% and standard deviation = 3.25% calculated using non-overlapping data

Technical Rationale for Documenting Attainment for 1% Non-attainment Dissolved Oxygen Criteria Values

State/District Co-Regulators
June 14, 2010 Conference Call
Attachment C2

Rich Batiuk and Gary Shenk

Technical Rationale for 1%

- **Two separate analyses**
  - Evaluation of evidence for 'residual' of 1.5 percent or less non-attainment in dissolved oxygen criteria values across large ranges of model simulated load reductions across multiple Bay segments and designated uses (Batiuk)
  - Analysis of changes in the sensitivity of dissolved oxygen criteria attainment to simulated load reductions (Shenk)

Residual of 1.5% or Less Analysis

- Keep in mind:
  - ‘Stoplight plots’ already account for unallowable exceedences
  - Stoplight plots also account for restoration variances adopted in states’ WQS regulations
  - Stoplight plots are the result of a comprehensive analysis system: model scenario output -> regression generation -> transformation of monitoring data -> criteria assessment fully consistent with states/DC WQS regulations

Residual of 1.5% or Less Analysis

- 21 designated use-segments with non-attainment values ranging from 0.0% to 1.5% (will round down to 1%) based on May 24, 2010 stoplight plot presented to WQGIT
  - Mainstem Bay, major river, small tributaries and embayment segments all represented
  - 11 open-water, 8 deep-water and 2 deep-channel designated uses

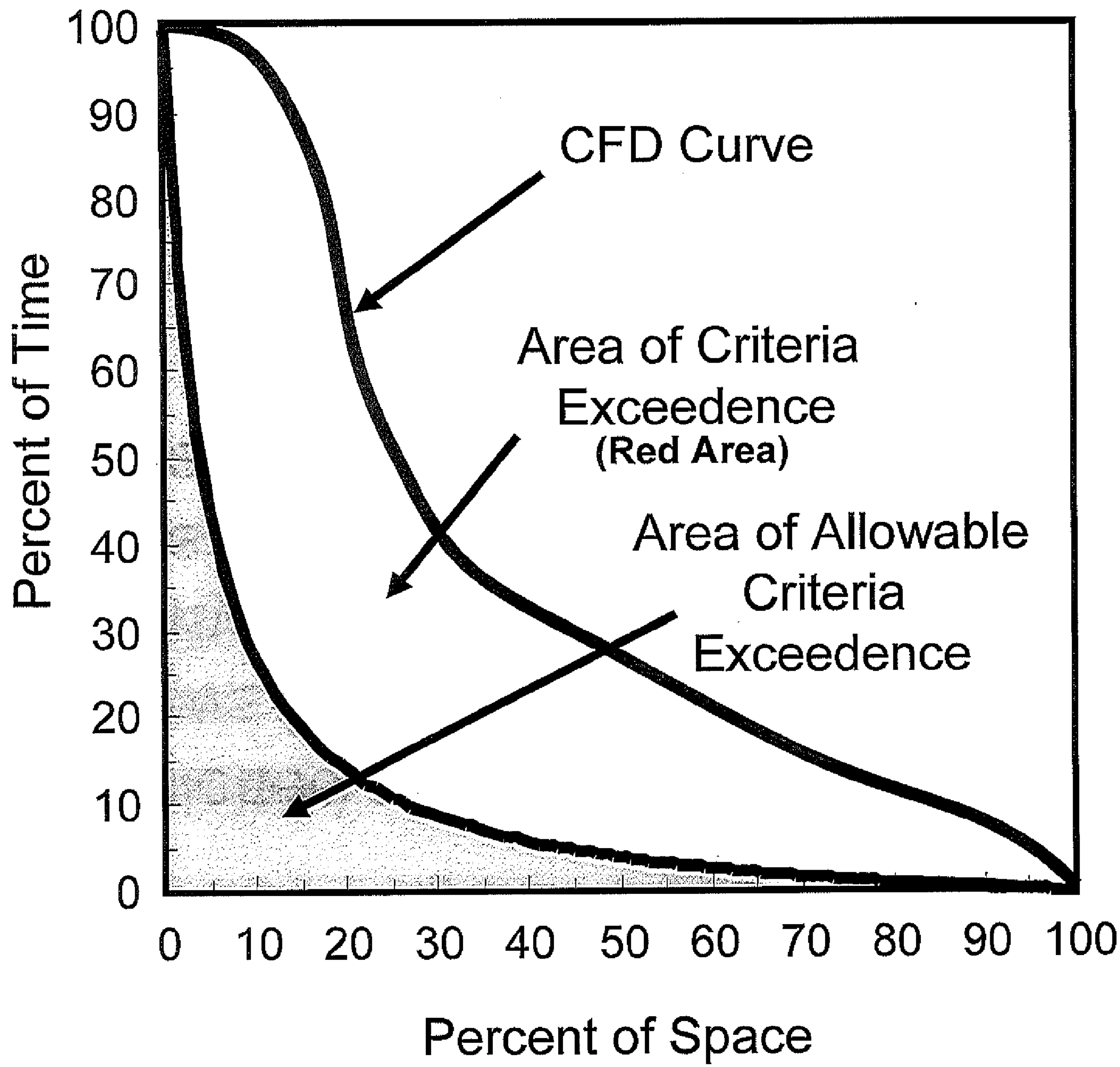
- Model simulated nitrogen load reductions ranged from 9 to 151 million pounds

Chesapeake Bay Segment	Designated Use	Criteria Non-attainment Range (%)	Model Simulated Nitrogen Load Range (million pounds/yr)
CB7	Open-water	0.5-0.0	200-141
CHOMH1	Open-water	0.1-0.0	254-179
CSHMH	Open-water	0.8-0.1	342-309
DCATF	Open-water	1.2-0.1	191-179
PAXTF	Open-water	1.0-0.6	190-179
DCPTF	Open-water	0.6-0.2	309-254
MAGMH	Open-water	1.3-0.3	342-191
MOBPH	Open-water	1.0-0.0	342-200
PIAMH	Open-water	0.1-0.1	191-179
TANMH	Open-water	1.5-0.1	342-309
YRKMH	Open-water	1.0-0.4	191-170
CB3MH	Deep-water	0.6-0.0	254-179
CB5MH	Deep-water	1.5-0.0	254-141
CHSMH	Deep-water	0.5-0.4	170-141
EASMH	Deep-water	0.8-0.2	200-170
MD5MH	Deep-water	1.5-0.1	191-141
MAGMH	Deep-water	0.5-0.5	170-141
PATMH	Deep-water	1.1-0.1	200-190
VA5MH	Deep-water	0.7-0.0	254-179
CB3MH	Deep-channel	0.2-0.1	200-190
EASMH	Deep-channel	1.3-0.0	190-170

Sensitivity to Load Reductions

- Plotting the change in unallowable exceedence (red area) per loading unit against the starting red area
  - Change in red area between two scenarios is divided by the change in the load
  - Changes in N and P loads are combined into a single measure:

Load units = (N + 10*P)/2


Percent of Time
100
90
80
70
60
50
40
30
20
10
0
CFD Curve
Area of Criteria Exceedence
(Red Area)
Area of Allowable Criteria Exceedence
0 10 20 30 40 50 60 70 80 90 100
Percent of Space

Sensitivity to Load Reductions

- When plotted against the starting red area, this allows for a direct comparison of sensitivity of the analysis system to load changes across different level of non-attainment

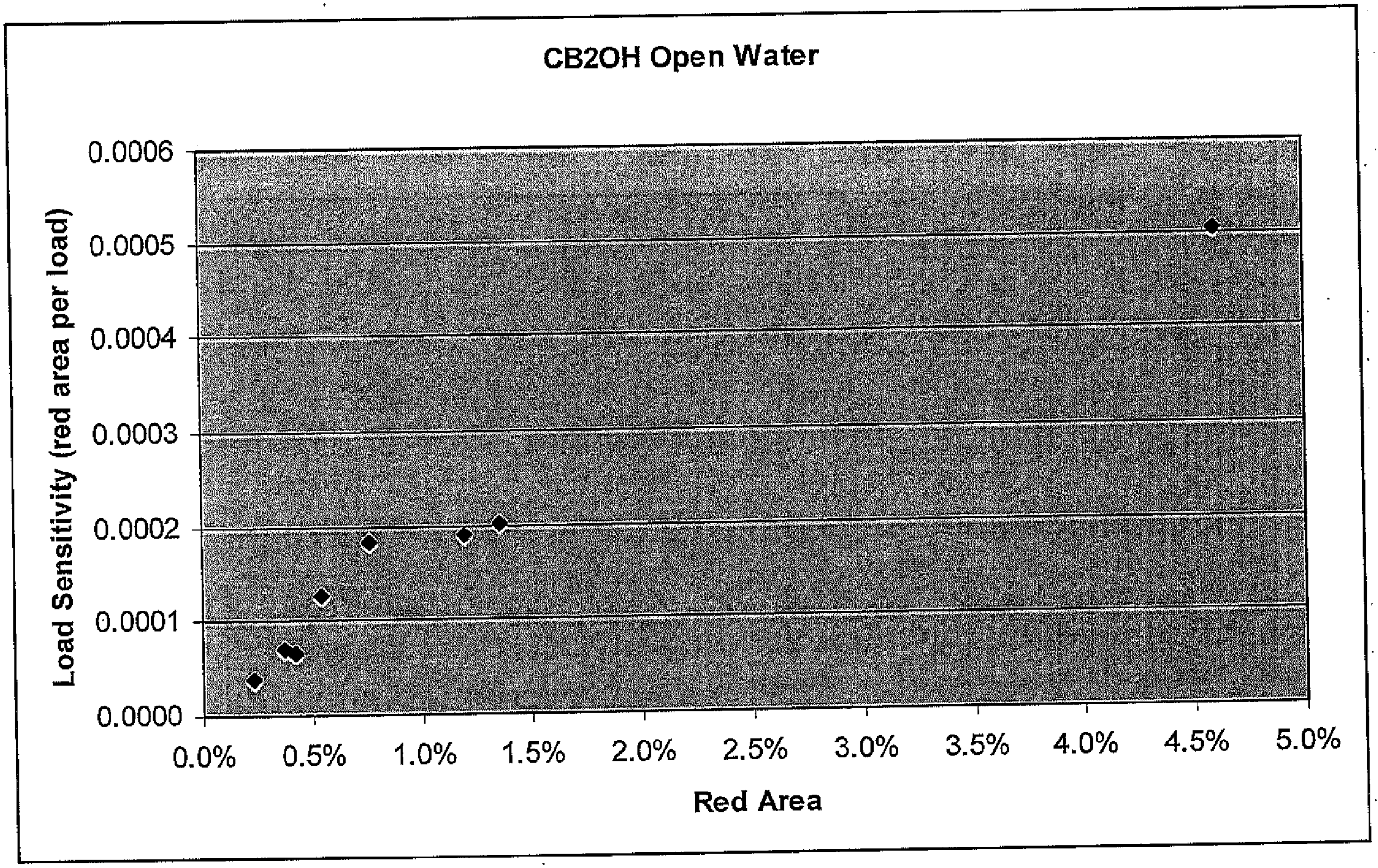
Sensitivity to Load Reductions

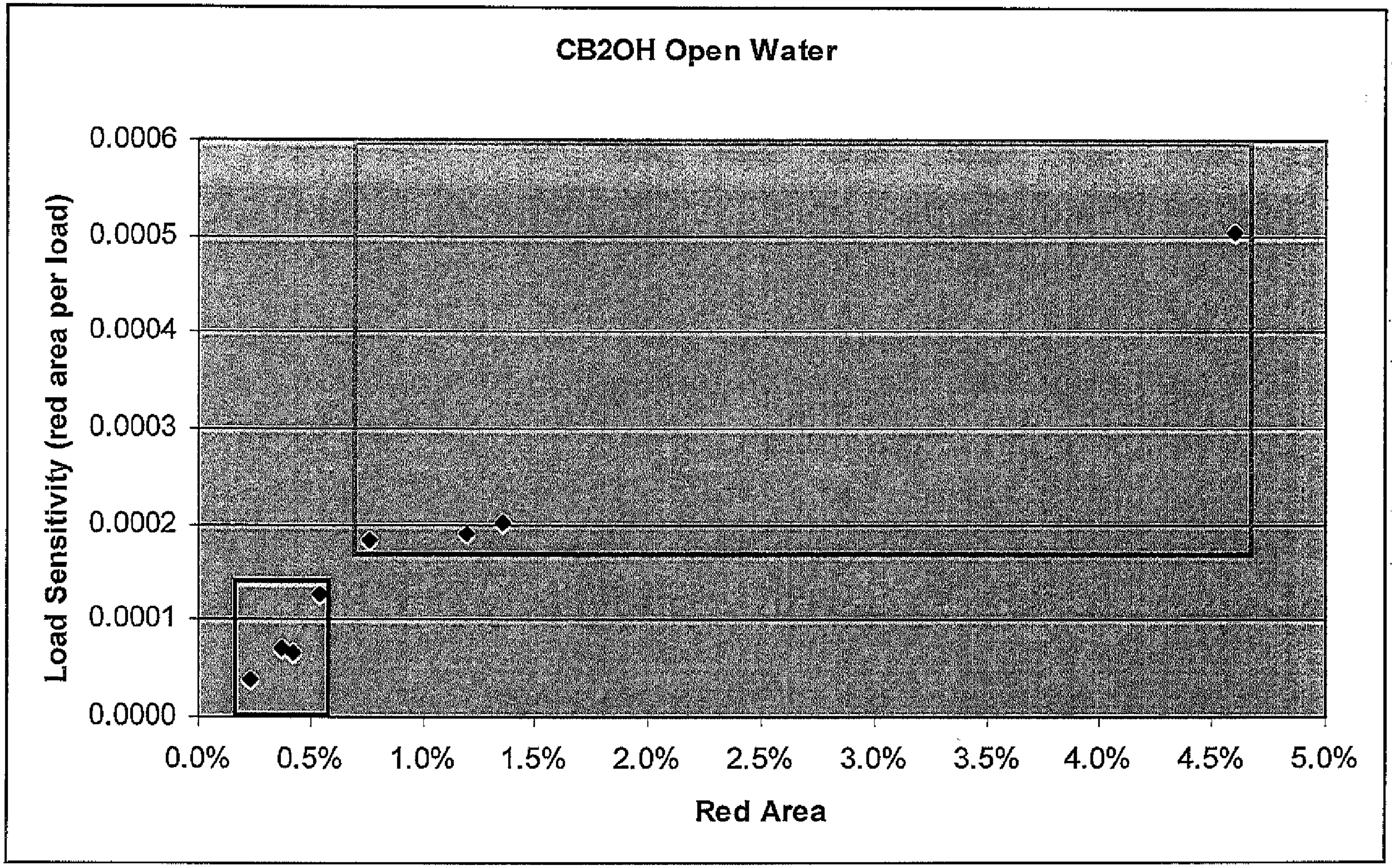
- 12 scenarios with eight 3-year periods for a total of 88 possible sensitivity assessments per designated use segment
  - Calculation involving scenarios where criteria were attained were not included in this analysis

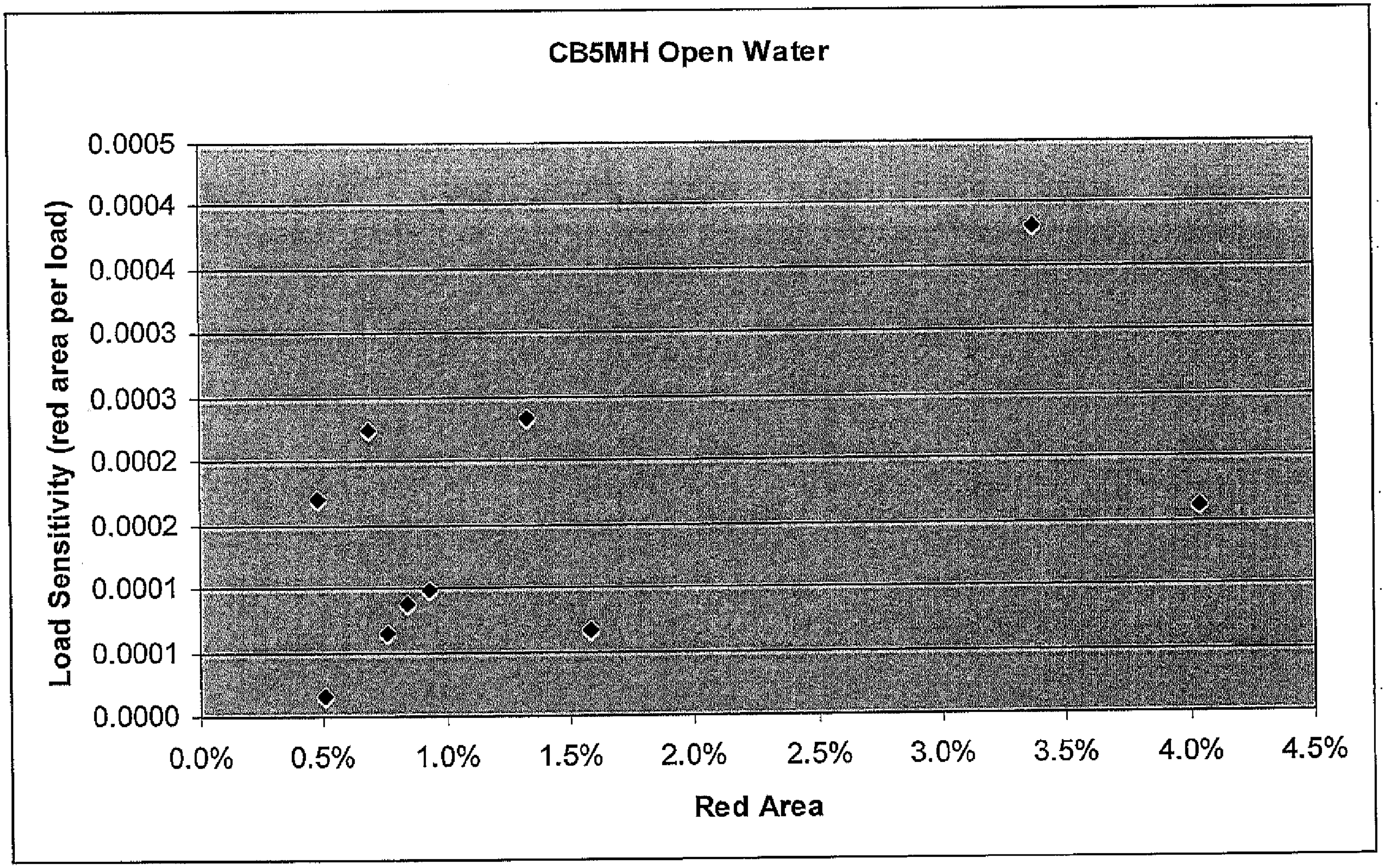
- Not amenable to tidal tributary segments
  - Loading are baywide, not specific to tributaries
  - Existing scenarios used for analysis have varying levels of reduction between different tributaries

Sensitivity to Load Reductions

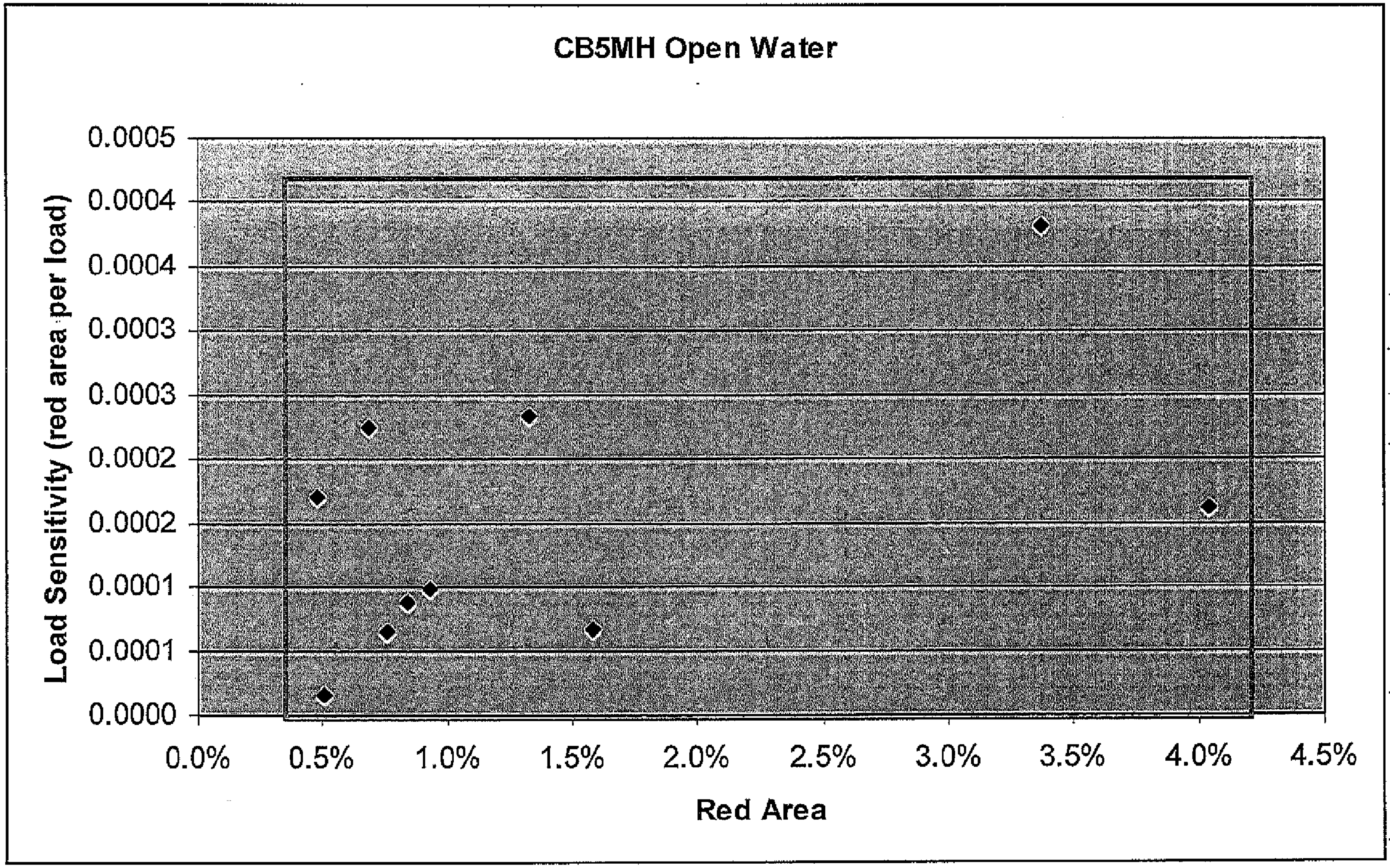
- Focused on segments driving the Bay TMDL
  - CB3MH, CB4MH, CB5MH for DW and DC and POTMH for DW
- Provided two open-water examples to show contrasts
  - CB2OH: drop in sensitivity at low non-attainment values
  - CB5MH: sensitivity to load reductions relatively constant throughout model simulated range

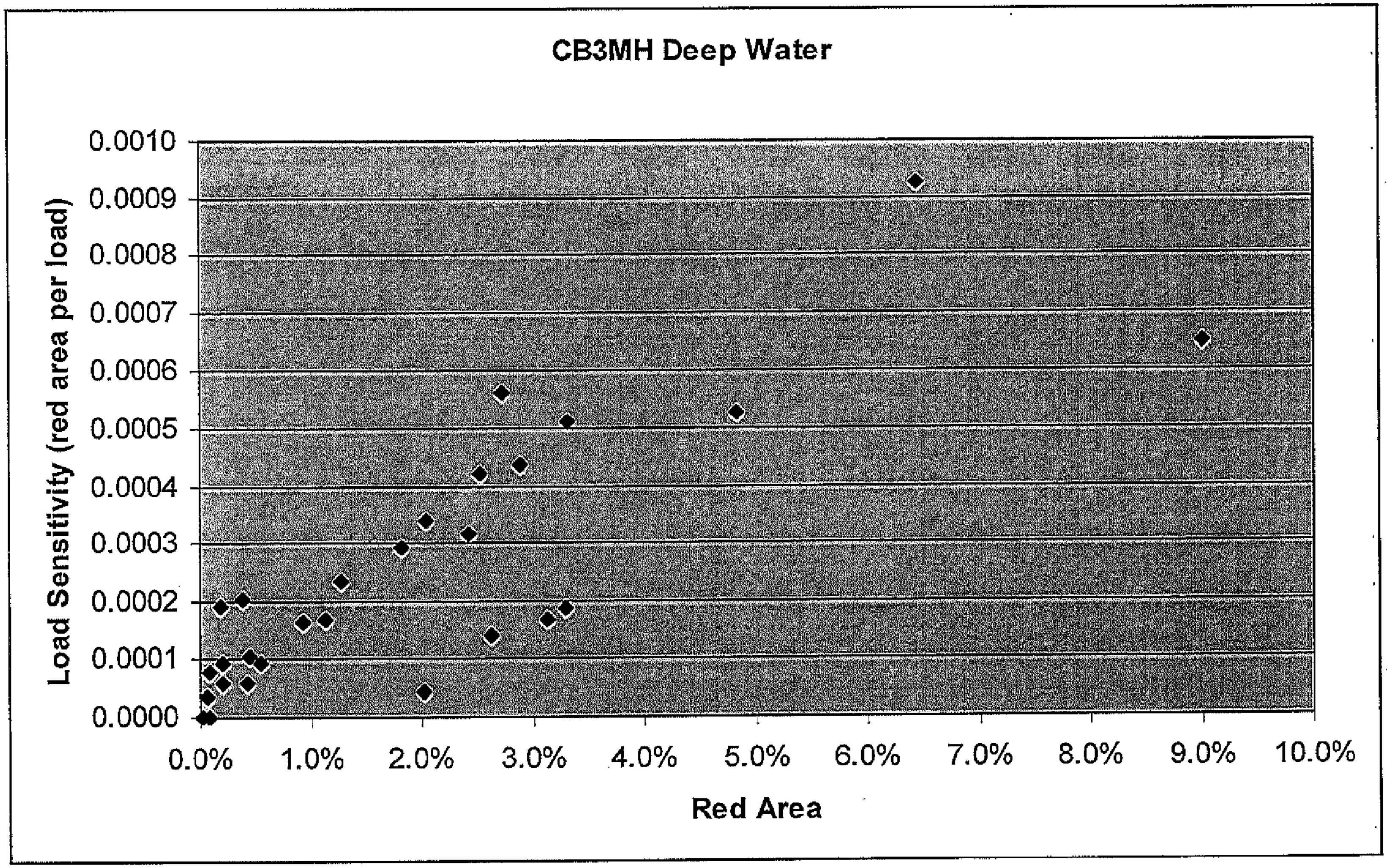

CB2OH Open Water
Load Sensitivity (red area per load)
0.0006
0.0005
0.0004
0.0003
0.0002
0.0001
0.0000
0.0%
0.5%
1.0%
1.5%
2.0%
2.5%
3.0%
3.5%
4.0%
4.5%
5.0%
Red Area

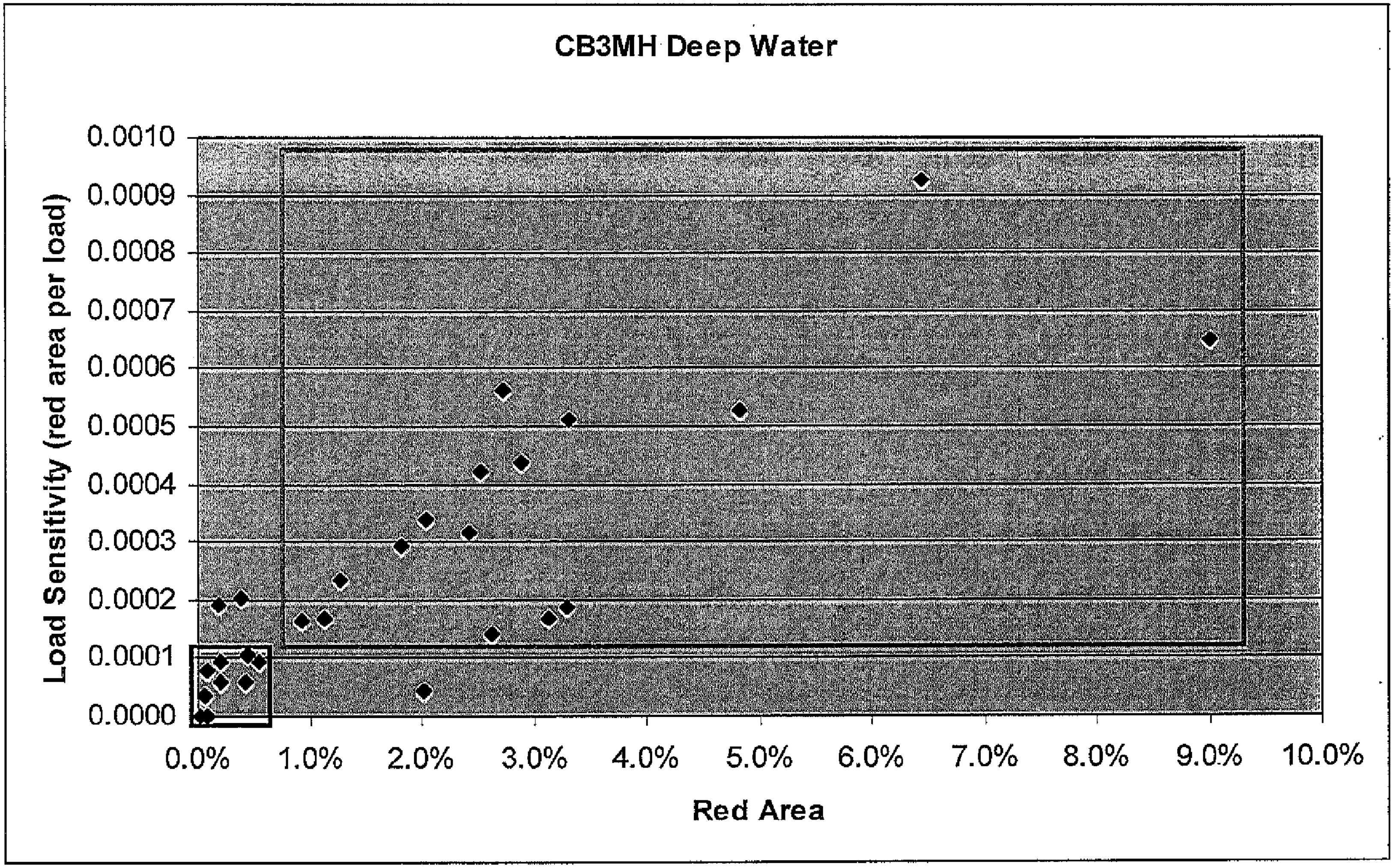

CB2OH Open Water
Load Sensitivity (red area per load)
0.0006
0.0005
0.0004
0.0003
0.0002
0.0001
0.0000
0.0%
0.5%
1.0%
1.5%
2.0%
2.5%
3.0%
3.5%
4.0%
4.5%
5.0%
Red Area

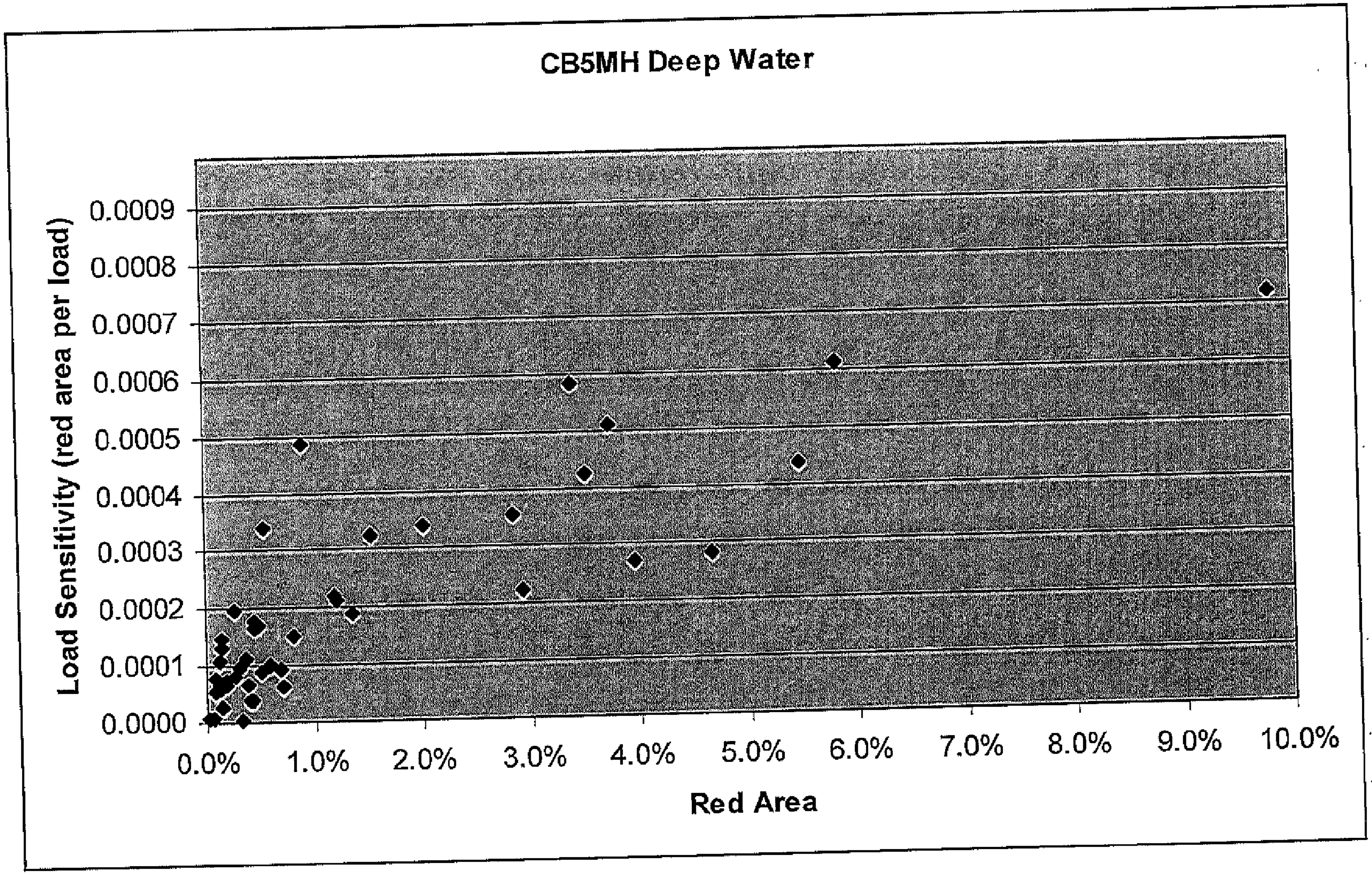


CB5MH Open Water
Load Sensitivity (red area per load)
0.0005
0.0004
0.0004
0.0003
0.0003
0.0002
0.0002
0.0001
0.0001
0.0000
0.0%
0.5%
1.0%
1.5%
2.0%
2.5%
3.0%
3.5%
4.0%
4.5%
Red Area

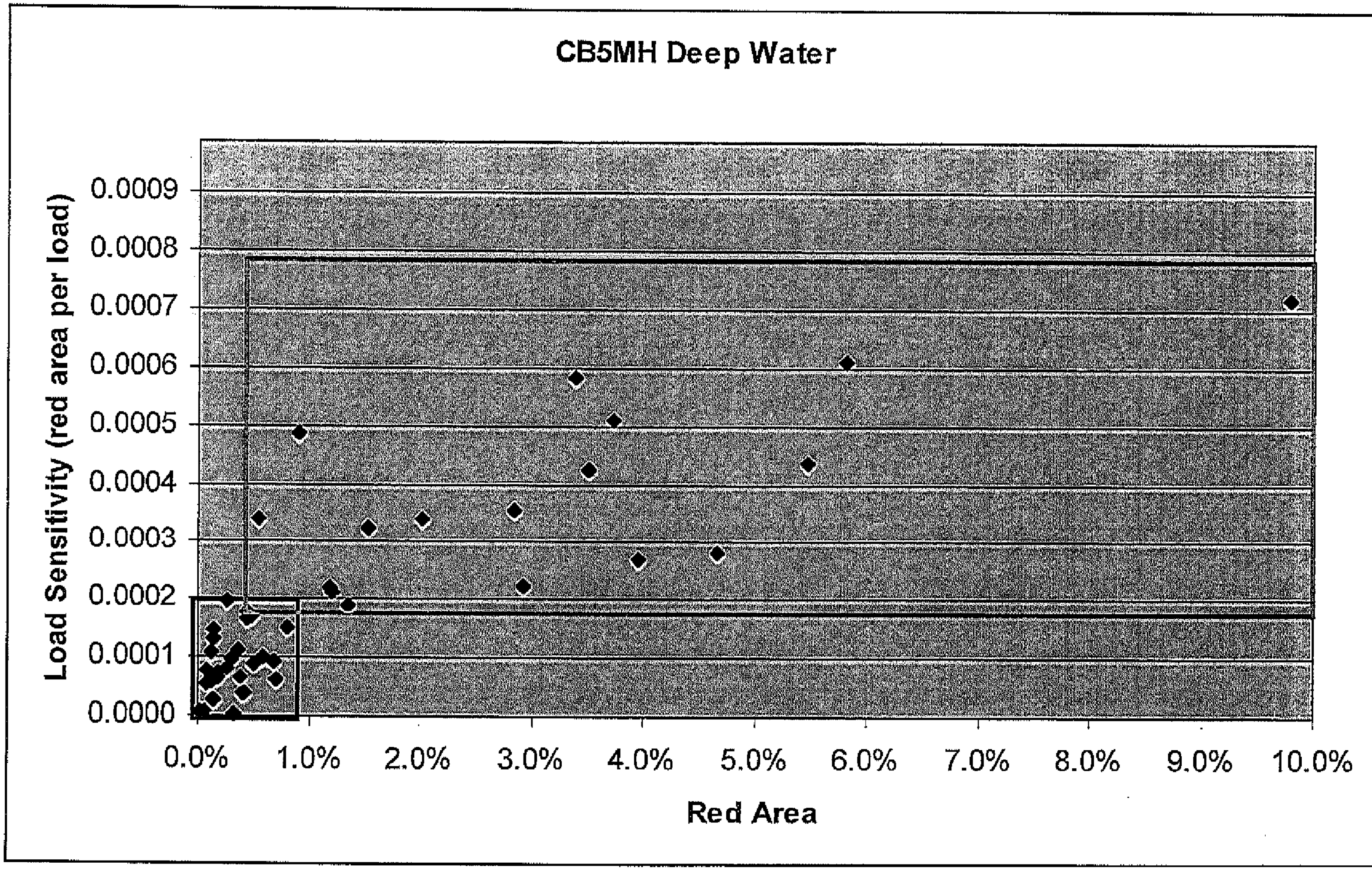

CB5MH Open Water
Load Sensitivity (red area per load)
0.0005
0.0004
0.0004
0.0003
0.0003
0.0002
0.0002
0.0001
0.0001
0.0000
0.0%
0.5%
1.0%
1.5%
2.0%
2.5%
3.0%
3.5%
4.0%
4.5%
Red Area

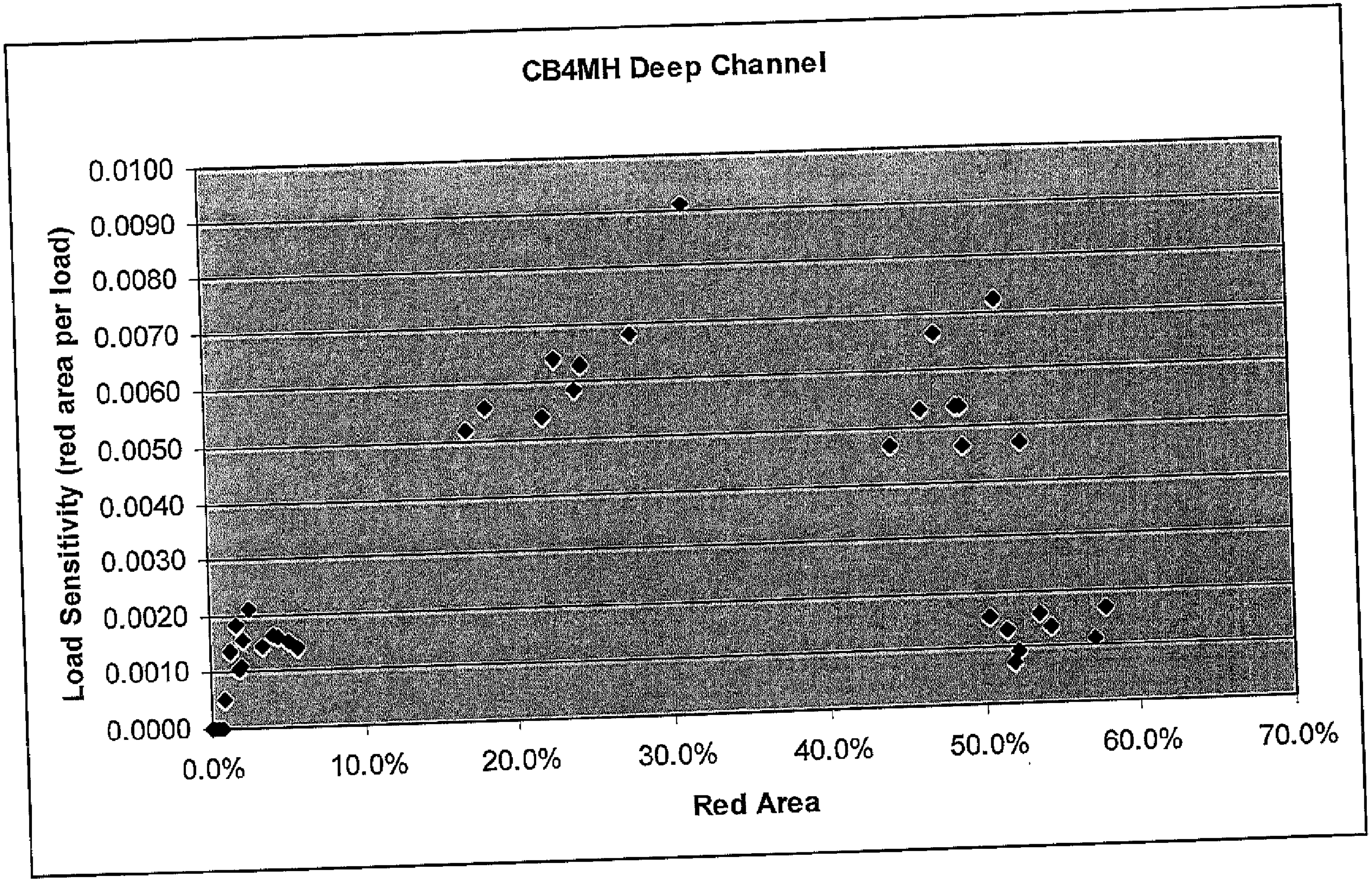

CB3MH Deep Water
Load Sensitivity (red area per load)
0.0010
0.0009
0.0008
0.0007
0.0006
0.0005
0.0004
0.0003
0.0002
0.0001
0.0000
0.0%
1.0%
2.0%
3.0%
4.0%
5.0%
6.0%
7.0%
8.0%
9.0%
10.0%
Red Area


CB3MH Deep Water
Load Sensitivity (red area per load)
0.0010
0.0009
0.0008
0.0007
0.0006
0.0005
0.0004
0.0003
0.0002
0.0001
0.0000
0.0%
1.0%
2.0%
3.0%
4.0%
5.0%
6.0%
7.0%
8.0%
9.0%
10.0%
Red Area

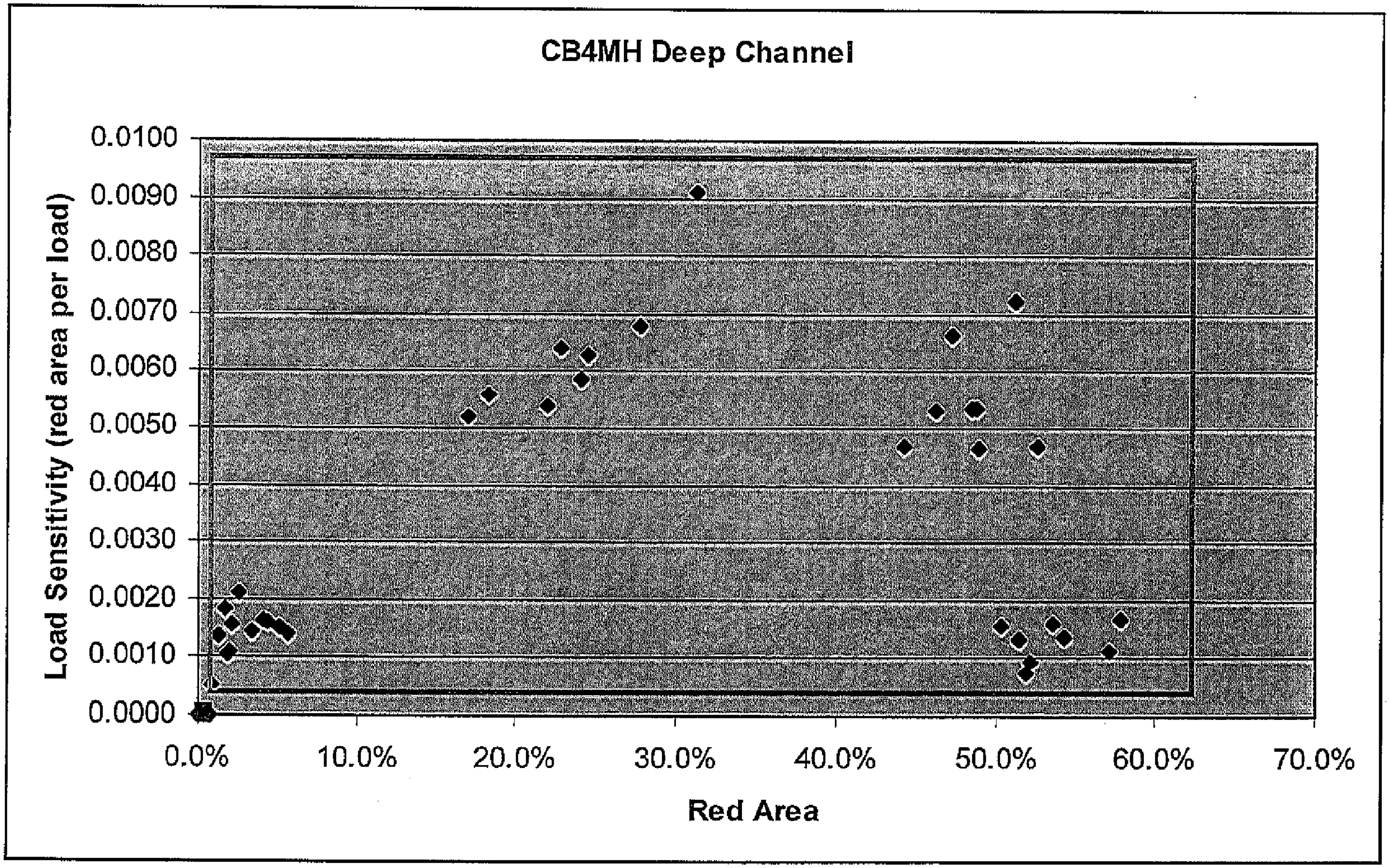


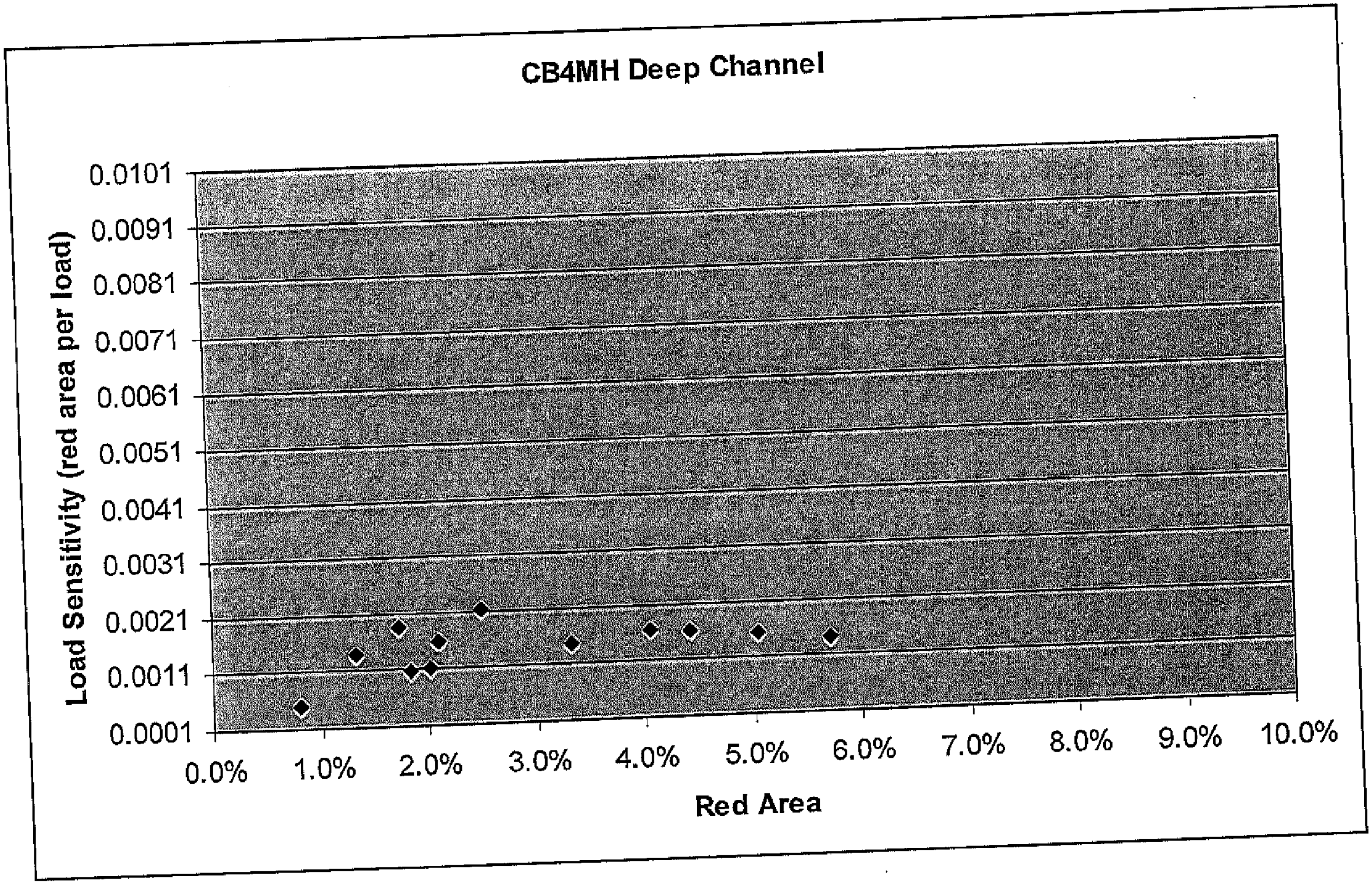
CB5MH Deep Water
Load Sensitivity (red area per load)
0.0009
0.0008
0.0007
0.0006
0.0005
0.0004
0.0003
0.0002
0.0001
0.0000
0.0%
1.0%
2.0%
3.0%
4.0%
5.0%
6.0%
7.0%
8.0%
9.0%
10.0%
Red Area


CB5MH Deep Water
Load Sensitivity (red area per load)
0.0009
0.0008
0.0007
0.0006
0.0005
0.0004
0.0003
0.0002
0.0001
0.0000
0.0%
1.0%
2.0%
3.0%
4.0%
5.0%
6.0%
7.0%
8.0%
9.0%
10.0%
Red Area

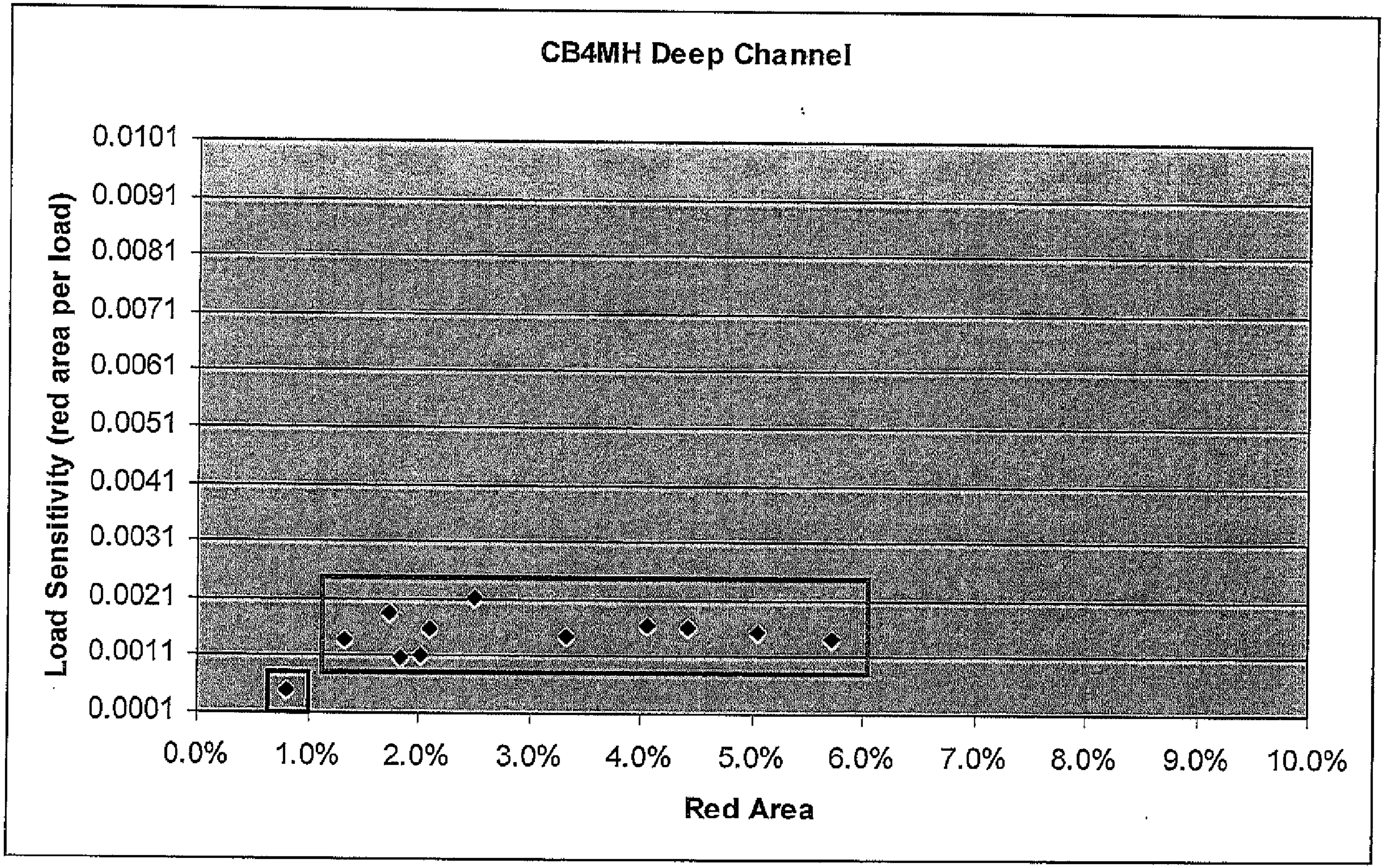


CB4MH Deep Channel
Load Sensitivity (red area per load)
0.0100
0.0090
0.0080
0.0070
0.0060
0.0050
0.0040
0.0030
0.0020
0.0010
0.0000
0.0%
10.0%
20.0%
30.0%
40.0%
50.0%
60.0%
70.0%
Red Area


CB4MH Deep Channel
Load Sensitivity (red area per load)
0.0100
0.0090
0.0080
0.0070
0.0060
0.0050
0.0040
0.0030
0.0020
0.0010
0.0000
0.0%
10.0%
20.0%
30.0%
40.0%
50.0%
60.0%
70.0%
Red Area



CB4MH Deep Channel
Load Sensitivity (red area per load)
0.0101
0.0091
0.0081
0.0071
0.0061
0.0051
0.0041
0.0031
0.0021
0.0011
0.0001
0.0%
1.0%
2.0%
3.0%
4.0%
5.0%
6.0%
7.0%
8.0%
9.0%
10.0%
Red Area


CB4MH Deep Channel
Load Sensitivity (red area per load)
0.0101
0.0091
0.0081
0.0071
0.0061
0.0051
0.0041
0.0031
0.0021
0.0011
0.0001
0.0%
1.0%
2.0%
3.0%
4.0%
5.0%
6.0%
7.0%
8.0%
9.0%
10.0%
Red Area

Chesapeake Bay Segment	Designated Use	Red Area with Low Sensitivity to Load Reductions (%)
CB3MH	Deep-water	0.2
CB4MH	Deep-water	0
CB5MH	Deep-water	1
POTMH	Deep-water	1
CB3MH	Deep-channel	1-1.5
CB4MH	Deep-channel	1
CB5MH	Deep-channel	N/A

Attainment at 1% Non-Attainment

- 21 designated use-segments with non-attainment values ranging from 0.0% to 1.5% across wide N load reductions

- Significant drop-off in sensitivity to load reductions documented at low non-attainment percentages
  - 1% being the consistent level at which sensitivity decreases across most of the designated use-segments driving the Bay TMDL

Attainment at 1% Non-Attainment

- The analysis system has been shown to be significantly less sensitive to the effects of further model simulated load reductions at and below the 1% non-attainment level

- These findings support documentation of attainment of designated use-segments with model simulated dissolved oxygen criteria non-attainment at 1%
  - No evidence documented for either analysis supporting a higher percentage (e.g., 2-3%)

ATTACHMENT F

MARYLAND ASSOCIATION OF MUNICIPAL WASTEWATER AGENCIES, INC.

VIRGINIA ASSOCIATION OF MUNICIPAL WASTEWATER AGENCIES, INC.

MEMORANDUM

TO: **CBP Water Quality Steering Committee Representatives**
CBP Nutrient Subcommittee Representatives
CBP Reevaluation Technical Workgroup Representatives

FROM: **V/MAMWA CBP Team**

CC: **MAMWA Board of Directors**
VAMWA Board of Directors

DATE: **January 21, 2009**

RE: **BMP Efficiencies**

Summary

This memorandum provides the recommendations of the Virginia and Maryland Associations of Municipal Wastewater Agencies ("V/MAMWA") on the appropriate treatment of BMP efficiencies for nutrients and sediment in the development of the Bay-wide TMDL. Recent model runs have predicted that attainment of Bay water quality standards will be much more difficult than indicated by previous models, posing serious questions of attainability. Reduced BMP efficiencies of the new model are one of the reasons for this discrepancy. Although V/MAMWA concur that model calibration scenarios should use "historical average" BMP efficiencies, we recommend that the Bay Program develop an alternate set of BMP efficiencies for implementation scenarios, reflecting improved BMP installation, operation, and maintenance. Such an approach would address the well-known need for such improvements, and also help address attainability concerns associated with the present TMDL process.

Background

Reduced BMP efficiencies are one of the reasons that the Phase 5 WSM predicts that attainment of water quality standards will be more difficult that predicted by previous model versions. In many cases, the modeled BMP efficiencies were reduced to be more conservative or realistic, reflecting the fact that many BMPs have historically not achieved intended design or research efficiencies. As stated in the Year 1 BMP Report entitled *Process for Developing BMP Definitions and Effectiveness Estimates*:

> Effectiveness recommendations should reflect operational conditions, defined as the average watershed wide condition…

The Year 1 BMPs (and presumably, the forthcoming Year 2 BMP report) cite numerous examples where modeled BMP efficiencies were made more conservative to reflect, in part, variability associated with BMP installation, operation, and maintenance. For example, the report on riparian buffers states the following:

> Based on discussions with researchers and literature reviews, a 20% reduction in the effectiveness values is applied to efficiencies from literature sources to account for spatial, temporal and management variability…

Similarly, the Year 1 report on urban wet ponds and wetlands states:

> The uncertainty in how improper maintenance will adjust BMP efficiencies supports the recommendation to use a more conservative percent removal estimate.

For some practices, the variability in efficiency is a function of controllable variables, such as tillage practice and planting date associated of cover crops. In other cases, BMP efficiencies were kept lower than literature values simply to be more conservative or for other programmatic reasons. For example, the Year 1 report on dry detention basins states:

> The CBP approved effectiveness estimates for Dry Detention Ponds/Basins and Hydrodynamic Structures were not changed based on the recommendation of the USWG. However, the function and actual effectiveness of these structures needs further evaluation since available literature does suggest somewhat higher removal rates…

In summary, many of BMP efficiencies in the Phase 5 WSM were either lowered or kept low to reflect both uncontrollable and controllable variability, including how the practices are installed, operated, or maintained. Although the model assumptions are intended to be realistic (to achieve the best model calibration) rather than explicitly conservative, it appears that at least some of BMP efficiency values have intentionally been set to conservatively low values.

Model Calibration v. TMDL/Tributary Strategy BMP Efficiencies

It is reasonable that model calibration scenarios should assume historical "average" management conditions. Any other approach—including the use of conservatively low values—would make the model less accurate and thus adversely impact model calibration. However, it is not necessary for forward-looking management scenarios to retain the assumption of historically-average BMP management. Rather, improvements in the way BMPs are installed, operated, and maintained are a viable implementation component. To state the concept another way, TMDL implementation recommendations should be based on the manner in which BMPs *should* be managed, not necessarily how they have historically been managed. This will allow the Bay tributary strategies to explicitly consider the well-documented need for improvements in BMP installation, operation, and maintenance.

One example of where the Bay Program and States have not assumed less than acceptable nutrient removal performance is for wastewater treatment plants. The performance expected and used in the model is based on properly installed, operated and maintained facilities. The standard for performance relative to design of any nutrient removal strategy (wastewater plants, BMPs, filter feeders, etc.) used in the Bay model should not be different.

Recommendations

Based on the discussion above, VAMWA and MAMWA make two related recommendations to the Chesapeake Bay Program:

1. All BMP efficiencies should be reviewed to ensure that the selected values used for model calibration are representative of *average* management conditions, not *conservatively low* estimates of management condition.
2. All BMP efficiencies used in future management/TMDL scenarios should be reviewed to ensure that the selected values used for model calibration are representative of BMPs that are installed, operated and maintained properly.

If you have any questions or would like to discuss these recommendations with V/MAMWA representatives, please contact Chris Pomeroy at (804) 716-9021 or chris@aqualaw.com.

* * *